



**HAL**  
open science

# A mixture of ellipsoidal densities for 3D data modelling

Denis Brazey, Antoine Godichon-Baggioni, Bruno Portier

► **To cite this version:**

Denis Brazey, Antoine Godichon-Baggioni, Bruno Portier. A mixture of ellipsoidal densities for 3D data modelling. 2023. hal-04209721

**HAL Id: hal-04209721**

**<https://hal.science/hal-04209721v1>**

Preprint submitted on 19 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A mixture of ellipsoidal densities for 3D data modelling

Denis Brazey\*, Antoine Godichon-Baggioni† and Bruno Portier ‡

## Abstract

In this paper, we propose a new ellipsoidal mixture model. This model is based on a new probability density function belonging to the family of elliptical distributions and designed to model points spread around an ellipsoidal surface. Then, we consider a mixture model based on this density, whose parameters are estimated with the help of an EM algorithm. The properties of the estimates are studied theoretically and empirically. The algorithm is compared to a state of the art ellipse fitting method and experimented on 3D data.

**Keywords:** Ellipsoidal mixture model; Elliptical distribution; EM algorithm.

## 1 Introduction

Fitting geometric primitives to a set of noisy points holds significant practical relevance across a multitude of scientific domains, including environmental science [5], agriculture [26], computer-aided design (CAD) [29], industrial applications [22], robotics [32], autonomous vehicles [7], and LiDAR applications [33].

In particular, geometric modeling offers a powerful means of representing objects present in images captured by imaging sensors. These sensors provide 2D or 3D point sets that describe the external surfaces of objects. Describing these data using simple geometric shapes such as spheres, cylinders, planes, and ellipsoids enables us to summarize and extract meaningful information. This high-level representation is crucial for efficient analysis and understanding of image content.

In practice, two fundamental scenarios arise: shape detection and shape fitting. In classical fitting problems, it is assumed that the majority of points belong to the shape of interest. Consequently, the entire data set is modeled with a single instance of the shape, with some points considered as outliers. The accuracy of this modeling is inevitably impacted by the noise in the data points [16], as well as their susceptibility to outliers and high levels of missing data. Typically, these algorithms are grounded in the principles of least squares, where the minimized distance is usually algebraic or geometric [14, 13, 1].

Conversely, in shape detection problems, the data may contain a high proportion of outliers that belong to other objects. The objective here is to extract one or multiple instances of a given shape without prior knowledge of which observations belong to them. Techniques like RANSAC and Hough transform are commonly employed to address this challenge. RANSAC randomly selects sets of inlier points, fits the primitive shape, and retains those closest to the data [37]. The Hough transform, on the other hand, employs a voting scheme in a discretized parameter space [34]. However, these

---

\*Société Prynel, RD974, 21190 Corpeau, France

†Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université, 75005 Paris, France

‡Laboratoire de Mathématiques de l'INSA, INSA Rouen-Normandie, 76800 Saint Etienne du Rouvray

methods often exhibit sensitivity to parameter tuning. An alternative approach is to first apply a segmentation algorithm, followed by a fitting algorithm on the detected clusters. Recently, deep neural networks have also been employed for primitive detection, segmenting point clouds into clusters and predicting classification and membership scores for subsequent fitting steps [27, 15, 31].

Mixture models [19] are statistical tools that can efficiently classify data and estimate a probabilistic model for each component. They find applications in various image processing tasks, including background subtraction [42], image segmentation [39], head detection [4], and ellipsoid fitting [41].

In this work, our focus lies in modeling a 3D point cloud using a mixture of ellipsoidal shapes. The primary challenge is selecting the component density within the mixture. We propose employing a parametric density of the form:

$$f_{\theta}(x) = C \exp\left(-\frac{d_{\theta}^2(x)}{2\sigma^2}\right), \quad x \in \mathbb{R}^d$$

where  $C$  represents the normalization constant,  $d_{\theta}(x)$  quantifies the signed distance between point  $x \in \mathbb{R}^d$  and the surface  $S$  defined by the parameter vector  $\theta$ , and  $\sigma$  controls the noise level. This distribution characterizes points distributed around the surface  $S$  of the considered shape with fluctuations in the normal direction. The surface is defined by the zero level line of the distance function  $d_{\theta}$ , where the density  $f$  reaches its maximum value.

To fit an ellipsoid with a center  $\mu \in \mathbb{R}^d$  and shape matrix  $\Sigma \in \mathbb{R}^{d \times d}$ , we propose using the following signed distance:

$$d_{\theta}(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} - 1,$$

resulting in a new pdf and leading to a new mixture model based on this density. Let us note that for the choice  $d_{\mu}(x) = \|x - \mu\|$ , where  $\|\cdot\|$  denotes the Euclidian norm in  $\mathbb{R}^d$ , then  $S = \{\mu\}$  and the density  $f$  reduces to the usual Gaussian density of parameters  $\mu$  and  $\sigma$ . When considering  $d_{(\mu,r)}(x) = \|x - \mu\| - r$ , then  $S$  is the sphere of center  $\mu \in \mathbb{R}^d$  and radius  $r > 0$  and we obtain the density introduced in [4].

The primary objectives of this study are twofold: first, to develop efficient methods for parameter estimation of this new density, providing direct estimates and their convergence rates; second, to propose iterative estimation methods with improved behavior, particularly relevant for estimating the parameters of the mixture model. Finally, we estimate the parameters of this mixture model using the EM algorithm [6], where the maximization step is obtained thanks to the aforementioned iterative estimates. Finally, this approach seems to be confirmed by experiments on real and simulated data.

The structure of the paper is as follows: Section 2 introduces the ellipsoidal mixture model, while Section 3 delves into the estimation of model parameters. Section 4 presents results obtained from simulated data, and Section 5 discusses experiments conducted on real data. The proofs and additional results are deferred to the Appendix.

## 2 The ellipsoidal mixture model

In this section, we introduce the new ellipsoidal density and establish some useful properties and results which will be used in the rest of the paper.

## 2.1 The new ellipsoidal density

In order to model points spread around an ellipsoidal surface, we propose to use a parametric density of the form

$$f_{\theta}(x) = C \exp\left(-\frac{d_{\theta}^2(x)}{2\sigma^2}\right), x \in \mathbb{R}^d \quad (2.1)$$

where  $C$  is the normalization constant,  $d_{\theta}(x)$  quantify the signed distance between the point  $x \in \mathbb{R}^d$  and the surface  $S$  of parameter vector  $\theta$  and  $\sigma$  controls the level of noise.

Let us discuss the choice of the distance function. A  $d$ -dimensional ellipsoid is characterized by a center  $\mu \in \mathbb{R}^d$  and a shape matrix  $\Sigma \in \mathbb{R}^{d \times d}$  symmetric and positive definite. The ellipsoidal surface  $S$  is defined as follows

$$S = \left\{ x \in \mathbb{R}^d, (x - \mu)^T \Sigma^{-1} (x - \mu) = 1 \right\}. \quad (2.2)$$

The center  $\mu$ , defining the location of the ellipsoid, is the intersection between the  $d$  axes of the ellipsoid. The shape matrix  $\Sigma$  defines the shape and the orientation of the surface. More precisely, we can consider the decomposition  $\Sigma = P D P^{-1}$ , where  $D$  is the diagonal matrix of eigenvalues  $(\lambda_1, \dots, \lambda_d)$  of  $\Sigma$  such that  $0 < \lambda_1 = \lambda_{\min}(\Sigma) \leq \dots \leq \lambda_d = \lambda_{\max}(\Sigma)$  and  $P$  is the orthogonal matrix whose columns are the associated eigenvectors  $(v_1, \dots, v_d)$ . Principal axes of the ellipsoid are given by  $(v_i)_{i=1}^d$  and their lengths by  $(2\sqrt{\lambda_i})_{i=1}^d$ . Since the Euclidian distance is difficult to write analytically in the case of the ellipsoid, we decide to use the Mahalanobis distance defined as

$$d_m(x, \mu, \Sigma) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}, \quad (2.3)$$

where  $x \in \mathbb{R}^d$  is the considered point and  $\theta = (\mu, \Sigma)$  is the parameter of the ellipsoid. From (2.2), the ellipsoid is defined by the contour line of value one. Note that in the particular case where  $\Sigma = I_d$ , the Mahalanobis distance coincides with the Euclidian distance. Finally, the general density (2.1) rewrites

$$f(x) = C_d \exp\left(-\frac{1}{2\sigma^2} \left(\sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} - 1\right)^2\right), \quad (2.4)$$

where  $\mu \in \mathbb{R}^d$ ,  $\Sigma \in \mathbb{R}^{d \times d}$  is symmetric and positive definite and  $\sigma^2 > 0$ . The parameter  $\sigma$  controls the dispersion of the observations around the surface. The normalization constant  $C_d$  equals to

$$C_d = \frac{\Gamma(d/2)}{2 \pi^{d/2} |\Sigma|^{1/2} J_{d-1}(\sigma)},$$

where  $\Gamma$  denotes the Gamma function, and for  $q \in \mathbb{N}$  and  $\alpha > 0$ ,

$$J_q(\alpha) = \int_0^{\infty} t^q \exp\left(-\frac{(t-1)^2}{2\alpha^2}\right) dt.$$

The detailed calculus are given in Appendix B. As this distribution is designed to model points spread around an ellipsoid, we only consider the case where  $\sigma$  is small enough. More precisely, we suppose that the value of  $\sigma$  is largely smaller than the length of the smallest half-axis of the ellipsoid, namely  $\sqrt{\lambda_{\min}(\Sigma)}$ . This assumption means that fluctuations do not interact with the other side of the surface. On top of that, if this condition is not verified, the density will model points located near the

center of the ellipsoid, which is not our objective. The ellipsoidal density of parameters  $\mu = (0, 0)^T$ ,  $\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$  and  $\sigma^2 = 0.1$  is represented Figure 1(a) and a sample from the distribution is given Figure 1(b). The sample illustrates that fluctuations are gaussians oriented in the normal direction of the surface.

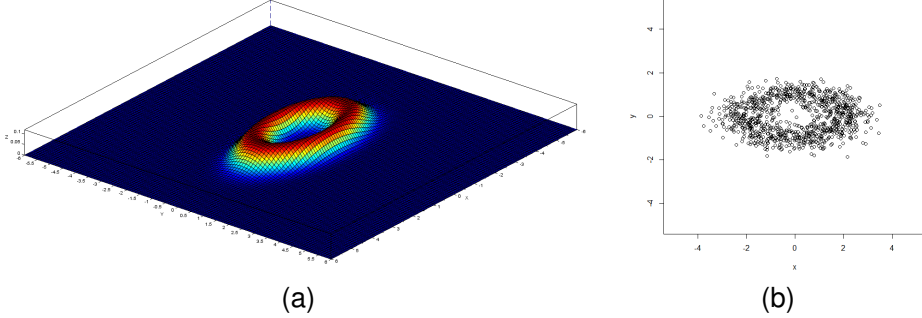


Figure 1: In (a) the ellipsoidal density in the case  $d = 2$  and in (b) a sample of size  $n = 1000$  from the distribution.

We now give some useful properties and results concerning the new distribution. Let us remark that in the case  $d \geq 2$ , our ellipsoidal distribution belongs to the family of elliptical distributions. Indeed, if we consider the function

$$g_\alpha(t) = \frac{\Gamma(d/2)}{2\pi^{d/2} J_{d-1}(\alpha)} \exp\left(-\frac{(\sqrt{t}-1)^2}{2\alpha^2}\right),$$

then the density  $f$  rewrites

$$f(x) = |\Sigma|^{-1/2} g_\sigma\left((x-\mu)^T \Sigma^{-1} (x-\mu)\right),$$

which means from Definition 1.5.2 from [21] that  $f$  is an elliptical pdf. Consequently, considering a random vector  $X$  of density  $f$  and  $Y = \Sigma^{-1/2} (X - \mu)$ , then  $Y$  has density  $g_\sigma(y^T y)$  and can be written under the form  $Y = WU$ , where  $W$  and  $U$  are independent,  $U$  is uniformly distributed on the  $d$ -dimensional unit sphere and  $W$  has the density

$$\varphi(t) = \frac{2\pi^{d/2}}{\Gamma(d/2)} t^{d-1}, \quad g_\sigma(t^2) = \frac{t^{d-1}}{J_{d-1}(\sigma)} \exp\left(-\frac{(t-1)^2}{2\sigma^2}\right) \mathbf{1}_{t \geq 0}.$$

The random vector  $X$  can therefore be written

$$X = \mu + \Sigma^{1/2} WU.$$

Let us now focus on the first moments of  $X$  and some other results which will be useful to prove some estimators properties in Section 3.

**Lemma 2.1.** *Let  $X$  be a random vector of  $\mathbb{R}^d$  (with  $d \geq 2$ ) of pdf  $f$  introduced in (2.4). For  $q \geq 1$ , we set  $J_q = J_q(\sigma)$ . Then,*

$$\mathbb{E}[X] = \mu \quad \text{and} \quad \mathbb{V}[X] = \mathbb{E}\left[(X-\mu)(X-\mu)^T\right] = \frac{J_{d+1}}{d J_{d-1}} \Sigma = d^{-1} \Sigma_*$$

where we set  $\Sigma_* = \frac{J_{d+1}}{J_{d-1}} \Sigma$ . In addition,

$$\mathbb{V} \left[ \sqrt{(X - \mu)^T \Sigma_*^{-1} (X - \mu)} \right] = \frac{J_{d+1}J_{d-1} - J_d^2}{J_{d+1}J_{d-1}} = \tilde{\sigma}^2.$$

The proof is given in Section A.1.

**Remark 2.1.** In the rest of the paper, we suppose that  $\sigma \leq 0.02$ , which is a reasonable value for our experimental settings (cf Appendix C). Results of the lemma are formulated for any dimension  $d \geq 2$  and for any value of  $\sigma$ . In the particular case where  $d = 3$  and  $\sigma$  is small enough, we can approximate  $J_2$  by  $C(1 + \sigma^2)$ ,  $J_3$  by  $C(1 + 3\sigma^2)$  and  $J_4$  by  $C(1 + 6\sigma^2 + 3\sigma^4)$ , where  $C = \sigma\sqrt{2\pi}$ .

Consequently, one can approximate  $\Sigma_* = (1 + 6\sigma^2 + 3\sigma^4)(1 + \sigma^2)^{-1}\Sigma$  and  $\tilde{\sigma}^2 = (1 + 3\sigma^4)(1 + 7\sigma^2 + 9\sigma^4 + 3\sigma^6)^{-1}\sigma^2$ . Finally, as  $\sigma$  is assumed sufficiently small, we deduce that  $\Sigma_* \simeq \Sigma$  and  $\tilde{\sigma}^2 \simeq \sigma^2$  (cf. Appendix B and Appendix C).

The following lemma gives results that will improve the estimation of parameter  $\mu$  via a Back-Fitting (BF for short) type algorithm.

**Lemma 2.2.** Let  $X$  be a random vector of  $\mathbb{R}^d$  (with  $d \geq 2$ ) of pdf  $f$  introduced in (2.4). Denoting

$$X^* := X - \frac{(X - \mu)}{\sqrt{(X - \mu)^T \Sigma_*^{-1} (X - \mu)}},$$

then

$$\mathbb{E}[X^*] = \mu \quad \text{and} \quad \mathbb{V}(X^*) = \mathbb{E} \left[ \left( W - \sqrt{\mathbb{E}[W^2]} \right)^2 \right] \frac{\Sigma}{d},$$

and  $\|\mathbb{V}[X^*]\| < \|\mathbb{V}[X]\|$ .

The proof is given in Section A.1. Lemma 2.1 and 2.2 combined with Remark 2.1 will be particularly useful for the estimation of the parameters of the distribution from a sample (see Section 3).

## 2.2 The finite mixture model

A finite mixture model is a probabilistic model characterised by a density function  $h$  of the form

$$h(x|\Theta) = \sum_{k=1}^K \pi_k f(x|\theta_k),$$

where  $f$  belongs to a parametric density family,  $K$  is the number of components of the mixture,  $(\pi_k)$  are the mixing weights satisfying  $\pi_k \geq 0$  and  $\sum_{k=1}^K \pi_k = 1$  and  $\Theta = (\pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K)$  is the vector containing all the parameters of the mixture.

The choice of the common density  $f$  depends on the data set. In many applications, Gaussian or Poisson densities have been used. In our case, since our data are 3-dimensional points spread around ellipsoidal surfaces, the pdf  $f$  is set to the new density introduced in (2.4) and the parameter vector  $\theta_k$  is then defined as  $\theta_k = (\mu_k, \Sigma_k, \sigma_k^2)$ .

### 3 Parameters estimation

This section concerns the estimation of the unknown parameters of the model from a sample in the case  $d = 3$  which is the major concern of the paper. Let us note that the estimation of the model in the case  $d = 2$  is straightforward. For shake of clarity, sections 3.1 and 3.2 are dedicated to the estimation of the parameters of the ellipsoidal density  $f$  while section 3.3 deals with the estimation of the ellipsoidal mixture model of density  $h$ .

#### 3.1 Direct estimation of the parameters

Let  $X_1, \dots, X_n$  be a sample of independent and identically distributed random vectors of  $\mathbb{R}^3$  with density  $f$  introduced in (2.4) with  $d = 3$  and under the assumption that  $\sigma \ll \sqrt{\lambda_{\min}(\Sigma)}$ . The objective is to estimate the parameters  $\mu, \Sigma$  and  $\sigma^2$  characterizing the distribution.

Let us start with the estimation of the center  $\mu$ . To this aim, we consider the empirical mean of the sample

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

It is clear from Theorem 2.1 that  $\bar{X}_n$  is an unbiased estimator of  $\mu$  and its rate of convergence derived from the law of the iterated logarithm, is given in Theorem 3.1. We now consider the estimation of the shape matrix  $\Sigma$ . We recall for this purpose the usual covariance matrix estimator

$$S_n = \frac{1}{n} \sum_{i=1}^n (X_i - \mu) (X_i - \mu)^T.$$

Thanks to the results of Theorem 2.1 and the strong law of large numbers, we can easily show that  $S_n$  is an unbiased and convergent estimator of  $J_4(\sigma) \Sigma / (3 J_2(\sigma))$ . When  $\sigma$  is sufficiently small, the ratio  $J_4 J_2^{-1}$  is close to 1 (see Remark 2.1) and in that case a reasonable estimator of  $\Sigma$  is  $\Sigma_n = 3 S_n$ . However, the parameter  $\mu$  being unknown, it is estimated by  $\bar{X}_n$  and we finally propose to estimate  $\Sigma$  by

$$\hat{\Sigma}_n = \frac{3}{n} \sum_{i=1}^n (X_i - \bar{X}_n) (X_i - \bar{X}_n)^T.$$

By using the decomposition

$$\hat{\Sigma}_n = \Sigma_n - 3 (\bar{X}_n - \mu) (\bar{X}_n - \mu)^T,$$

we can show that  $\hat{\Sigma}_n$  converges almost surely towards  $\Sigma^* = J_4(\sigma) J_2^{-1}(\sigma) \Sigma$  with rate (see Theorem 3.1) and consequently when  $\sigma$  is small,  $\hat{\Sigma}_n$  is a reasonable estimator of  $\Sigma$ .

We now focus on the estimation of parameter  $\sigma^2$ . In order to estimate this parameter, we introduce the estimator

$$\sigma_n^2 = \frac{1}{n} \sum_{j=1}^n (\xi_j - \mathbb{E}[\xi_j])^2$$

where for  $j = 1, \dots, n$ ,  $\xi_j = \sqrt{(X_j - \mu)^T (\Sigma^*)^{-1} (X_j - \mu)}$ . From the strong law of large numbers and from results of Theorem 2.1, we can deduce that  $\sigma_n^2$  converges almost surely towards  $\tilde{\sigma}^2 \simeq$

$\sigma^2$  when  $\sigma$  is sufficiently small (Remark 2.1). However, this estimator depends on the unknown parameters  $\mu$  and  $\Sigma^*$ . We therefore propose to estimate  $\sigma^2$  by

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{j=1}^n \widehat{\xi}_j^2 - \left( \frac{1}{n} \sum_{j=1}^n \widehat{\xi}_j \right)^2$$

where for  $j = 1, \dots, n$ ,  $\widehat{\xi}_j = \sqrt{(X_j - \bar{X}_n)^T \widehat{\Sigma}_n^{-1} (X_j - \bar{X}_n)}$ . Convergence results are given in the following theorem.

**Theorem 3.1.** *Let  $X_1, \dots, X_n$  be a sample of independent and identically distributed random vectors of  $\mathbb{R}^3$  with pdf  $f$ . Then, we have*

$$\|\bar{X}_n - \mu\| = O(a_n), \quad \|\widehat{\Sigma}_n - \Sigma_*\| = O(a_n) \quad \text{and} \quad |\widehat{\sigma}_n^2 - \sigma^2| = O(a_n) \quad \text{a.s.}$$

where  $a_n = \sqrt{(\log \log n)/n}$ .

The proof is given in Appendix A.2.

### 3.2 Iterative estimation of the parameters

In this section, we propose a more efficient estimation strategy, which will be of particular interest for estimating the parameters of mixture models. The estimate  $\bar{X}_n$  is an unbiased and convergent estimator of  $\mu$ . However, this estimator has a much larger variance than another unbiased and convergent estimator of  $\mu$  denoted as  $\bar{X}_n^*$  and defined by

$$\bar{X}_n^* = \bar{X}_n - \frac{1}{n} \sum_{i=1}^n \frac{(X_i - \mu)}{\sqrt{(X_i - \mu)^T \Sigma_*^{-1} (X_i - \mu)}}.$$

Indeed, using the results of Theorem 2.2, we can show that  $\bar{X}_n^*$  is an unbiased and convergent estimator of  $\mu$  satisfying  $\|\mathbb{V}(\bar{X}_n^*)\| < \|\mathbb{V}(\bar{X}_n)\|$ . Nevertheless, the estimator  $\bar{X}_n^*$  depends on unknown parameters  $\mu$  and  $\Sigma_*$  which must be estimated. We therefore propose the following strategy to improve the estimation of parameters  $\mu, \Sigma$  and  $\sigma^2$ . First, we estimate  $\mu$  by  $\widehat{\mu}_n^{(0)} = \bar{X}_n$ , then  $\Sigma_*$  by  $\widehat{\Sigma}_n^{(0)} = \widehat{\Sigma}_n$ . Then, we improve the estimation of  $\mu$  by considering

$$\widehat{\mu}_n^{(1)} = \bar{X}_n - \frac{1}{n} \sum_{i=1}^n \frac{(X_i - \widehat{\mu}_n^{(0)})}{\sqrt{(X_i - \widehat{\mu}_n^{(0)})^T (\widehat{\Sigma}_n^{(0)})^{-1} (X_i - \widehat{\mu}_n^{(0)})}} \quad (3.1)$$

and the estimation of  $\Sigma_*$  by considering

$$\widehat{\Sigma}_n^{(1)} = \frac{3}{n} \sum_{i=1}^n (X_i - \widehat{\mu}_n^{(1)}) (X_i - \widehat{\mu}_n^{(1)})^T. \quad (3.2)$$

The rate of convergence of these new estimates can be derived from Theorem 3.1. In addition, the process (3.1)-(3.2) can be iterated until convergence to improve the estimations of  $\mu$  and  $\Sigma_*$ . At the end of each iteration, we update the estimates as follows :  $\widehat{\mu}_n^{(0)} = \widehat{\mu}_n^{(1)}$  and  $\widehat{\Sigma}_n^{(0)} = \widehat{\Sigma}_n^{(1)}$ .



Although we did not succeed in proving that the backfitting estimates are theoretically better, this seems to be confirmed in practice (see Section 4). Finally, the parameter  $\sigma^2$  can then be estimated by

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{j=1}^n \widehat{\zeta}_j^2 - \left( \frac{1}{n} \sum_{j=1}^n \widehat{\zeta}_j \right)^2$$

with  $\widehat{\zeta}_i = \sqrt{(X_i - \widehat{\mu}_n^{(1)})^T (\widehat{\Sigma}_n^{(1)})^{-1} (X_i - \widehat{\mu}_n^{(1)})}$ .

### 3.3 Estimation of the mixture model parameters

This section concerns the estimation of a mixture model with  $K > 1$  components where the component  $k$  is described by the density

$$f(x | \theta_k) = \frac{(2\pi)^{-3/2} |\Sigma_k|^{-1/2}}{2\sigma_k(1 + \sigma_k^2)} \exp \left( -\frac{1}{2\sigma_k^2} (\sqrt{(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)} - 1)^2 \right)$$

where  $\theta_k = (\mu_k, \Sigma_k, \sigma_k)$  and  $\sigma_k$  is supposed sufficiently small. Let us note that the approximated version of  $C_3$  is used instead of its exact expression.

Let  $\mathbf{x} = (x_1, \dots, x_n)$  be a sample of points, where each  $x_i \in \mathbb{R}^3$  belongs to one component of the mixture. Given the sample  $\mathbf{x}$ , we can estimate the parameter vector  $\Theta$  of the mixture by using the Maximum-Likelihood estimator  $\widehat{\Theta}$  given by

$$\widehat{\Theta} = \arg \max_{\Theta} L(\mathbf{x}; \Theta) \quad (3.3)$$

where  $L$  denotes the log-likelihood function defined as follows

$$L(\mathbf{x}; \Theta) = \sum_{i=1}^n \log h(x_i | \Theta) = \sum_{i=1}^n \log \left( \sum_{k=1}^K \pi_k f(x_i | \theta_k) \right) \quad (3.4)$$

The maximization problem (3.3) is difficult to solve because of the non-linearity of the log-likelihood  $L$  in many parameters. The Expectation-Maximization (EM) algorithm has been introduced in [6] to solve this problem in the case of incomplete data (some extensions can be found in [20]).

The main idea of EM is to introduce a hidden variable  $z_i = (z_{i1}, \dots, z_{iK})$  given the membership of each  $x_i$  to the clusters  $(C_1, \dots, C_K)$ . The hidden variable is defined by  $z_{ik} = \mathbb{1}_{\{x_i \in C_k\}}$  and each possible value leads to a clustering of data. Data points  $\mathbf{x}$  are called *observed data* while  $\mathbf{z} = (z_1, \dots, z_n)$  are called *missing data* since these values are not observed. We can then introduce  $\mathbf{z}$  in (3.4) to obtain the so-called complete log-likelihood

$$L_c(\mathbf{x}, \mathbf{z}; \Theta) = \sum_{i=1}^n \sum_{k=1}^K z_{ik} (\log(\pi_k) + \log(f(x_i | \theta_k))) \quad (3.5)$$

The EM algorithm consists in maximizing the conditional expectation of the complete log-likelihood (3.5) given the observation  $\mathbf{x}$  and the current parameters estimate  $\Theta^{(j)}$  at iteration  $j$ . The quantity  $t_{i\ell} = \mathbb{E} [z_{i\ell} | \mathbf{x}, \Theta^{(j)}]$  is called membership probability of point  $x_i$  to cluster  $C_\ell$  and can be estimated with Bayes rule. Then, the EM algorithm consists of iterating the following two steps until convergence

- **E-step** : Compute the membership probabilities  $t_{i\ell}$  given by

$$t_{i\ell} = \frac{\pi_\ell^{(j)} f(x_i | \theta_\ell^{(j)})}{\sum_{k=1}^K \pi_k^{(j)} f(x_i | \theta_k^{(j)})} \quad (3.6)$$

using the current parameters estimate  $\Theta^{(j)} = (\pi_1^{(j)}, \dots, \pi_K^{(j)}, \theta_1^{(j)}, \dots, \theta_K^{(j)})$ .

- **M-step** : Find parameters  $\Theta^{(j+1)}$  that maximize the conditional expectation of  $L_c$  given the observation  $\mathbf{x}$  and the current parameters estimate  $\Theta^{(j)}$

$$Q(\Theta) = \sum_{i=1}^n \sum_{k=1}^K t_{ik} (\log(\pi_k) + \log(f(x_i | \theta_k))) \quad (3.7)$$

To start the iterative algorithm, an initial value  $\widehat{\Theta}^{(0)}$  must be determined, for example using the  $K$ -means clustering algorithm (see [3] for example).

The maximization of  $Q$  in the M-step is performed analytically by computing partial derivatives with respect to each parameter and involves the assumption on parameter  $\sigma$  (for more details, see appendix D). For each component  $\ell \in [1, K]$  of the mixture, we obtain the following equations defining the estimators

$$\left\{ \begin{array}{l} \widehat{\mu}_\ell = \frac{1}{\sum_{i=1}^n t_{i\ell}} \left[ \sum_{i=1}^n t_{i\ell} X_i - \sum_{i=1}^n \frac{t_{i\ell} (X_i - \widehat{\mu}_\ell)}{\xi_{i,\ell}} \right] \\ \widehat{\Sigma}_\ell = \frac{3}{\sum_{i=1}^n t_{i\ell}} \sum_{i=1}^n t_{i\ell} (X_i - \widehat{\mu}_\ell) (X_i - \widehat{\mu}_\ell)^T \\ \widehat{\sigma}_\ell^2 = \frac{1}{\sum_{i=1}^n t_{i\ell}} \sum_{i=1}^n t_{i\ell} (\xi_{i,\ell} - \bar{\xi}_\ell)^2 \\ \widehat{\pi}_\ell = \frac{1}{n} \sum_{i=1}^n t_{i\ell} \end{array} \right. \quad (3.8)$$

with

$$\xi_{i,\ell} = \sqrt{(X_i - \widehat{\mu}_\ell)^T \widehat{\Sigma}_\ell^{-1} (X_i - \widehat{\mu}_\ell)} \quad \text{and} \quad \bar{\xi}_\ell = \frac{1}{n} \sum_{i=1}^n \xi_{i,\ell}.$$

In addition, since estimators  $\mu$  and  $\Sigma$  depends to each other, estimates of  $\mu$  and  $\Sigma$  are computed using the iterative backfitting procedure defined by (3.1)-(3.2). The EM algorithm and the iterative backfitting procedure have a similar stopping criterion. Algorithms are stopped when the difference between consecutive estimated parameters values is small or when a maximal number of iterations has been reached. The behaviour of the estimation algorithm will be studied on simulated data in the following section.

## 4 Experiments on simulated data

The objective of the experiments carried out in this section is to show that the proposed estimation methods (backfitting BF and EM) behave as expected in simulations. Parameters of the simulated

distributions have been chosen such that the assumption on parameter  $\sigma$  is satisfied. Samples of a random variable  $X$  of density  $f$  have been simulated in dimensions  $d = 2$  and  $d = 3$  using the decomposition  $X = \mu + \Sigma^{1/2} W U$ , a rejection sampling method on the random variable  $W$  and an inverse transform method on the random vector  $U$ . The case  $d = 2$  has been considered to compare the estimation method to a standard algorithm. More precisely, the ellipse estimation method BF is compared to [10] implemented in the open source OpenCV library [25]. Numerical results will be analyzed in the case of a single ellipsoid in section 4.1 and for a mixture model in section 4.2.

#### 4.1 The case of the single ellipsoid

The first set of experiments deals with the modeling of a single ellipsoid. We first consider random vectors  $X_d$ , with  $d = 3$  with ellipsoidal density  $f$  introduced in (2.4). For each experiment, we simulate  $N = 200$  independant samples of realizations of the random vector  $X_d$ . The parameters

$$\mu_{(2)} = (0,0)^T, \Sigma_{(2)} = \begin{pmatrix} 100^2 & 0 \\ 0 & 50^2 \end{pmatrix}, \sigma_{(2)} = 0.01 \text{ for } d = 2 \text{ and } \mu_{(3)} = (0,0,0)^T, \\ \Sigma_{(3)} = \begin{pmatrix} 100^2 & 0 & 0 \\ 0 & 50^2 & 0 \\ 0 & 0 & 50^2 \end{pmatrix}, \sigma_{(3)} = 0.01 \text{ for } d = 3.$$

Figure 2 shows boxplots describing the  $N$  estimations of parameters  $\mu_{(3)}$ ,  $\Sigma_{(3)}$  and  $\sigma_{(3)}$  obtained with the BF algorithm for different sample sizes. For the shape matrix, we compute the error  $E(\Sigma) = \left( \sum_{i,j=1}^d (\hat{\Sigma}_{i,j} - \Sigma_{i,j})^2 \right)^{1/2}$ . We note that the estimated values are close to the expected ones for a sufficiently high number of observations. It is clear that when the sample size increases, the accuracy of the results is improved and the variability decreases. Let us note that the same remarks can be made for the results obtained in the case  $d = 2$ .

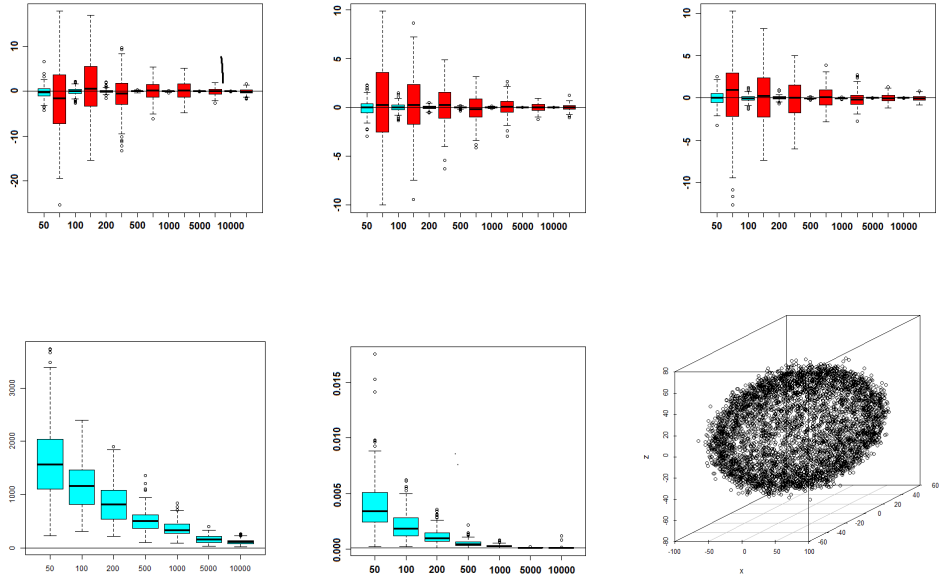


Figure 2: On the first row, boxplots of estimates of  $\mu_x$ ,  $\mu_y$  and  $\mu_z$ . On the second row, boxplots of  $E(\Sigma)$ , estimates of  $\sigma$  and a sample of size  $n = 1000$ .

In section 3.2, we mentioned that the variance of the estimator  $\overline{X}_n^*$  is smaller than the variance of the empirical mean  $\overline{X}_n$ . A comparison between these two estimators is presented on Figure 3 for  $d = 3$ . We observe that the variability of the center estimates is widely smaller with BF than with the empirical mean.

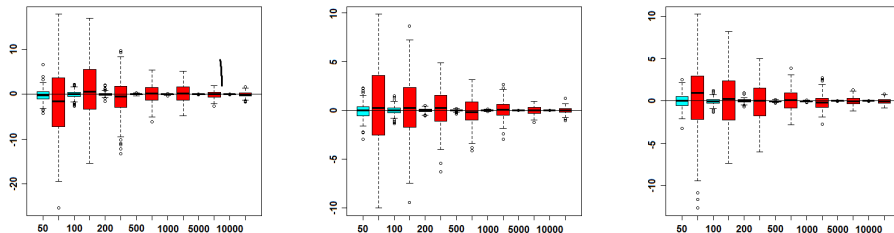


Figure 3: Boxplot of estimates of parameters  $\mu_x$ ,  $\mu_y$  and  $\mu_z$  for different sample sizes  $n$ . The BF method is plotted in blue and the empirical mean in red.

Finally, we compare the proposed estimation method to the ellipse fitting algorithm [10] implemented in the open source OpenCV library [25]. This reference method is widely used in computer vision to fit an ellipse to a set of image edges. We simulate samples from the density (2.4) with  $d = 2$  and with parameters  $\mu = (0, 0)$ ,  $\sigma = 0.01$  and  $\Sigma = \begin{pmatrix} 100^2 & 50^2 \\ 50^2 & 50^2 \end{pmatrix}$ . Obtained results are plotted on Figure 4. For the centre estimation, we observe that the proposed method offers estimations as accurate as the reference one. Moreover, the variability of the estimates is of the same order of magnitude for the two compared algorithms. For the shape matrix estimation, we note that the BF algorithm offers slightly better estimates than the reference one. However, the variability of the estimates computed with the BF method is larger.

## 4.2 A mixture model with three components

The aim of this second set of experiments is to show that the EM algorithm is able to separate and model several overlapping ellipsoids. The EM algorithm is initialized with a parameter vector  $\Theta^{(0)}$  computed from a classification obtained with the K-means clustering algorithm. We consider an ellipsoidal mixture model made up of 3 overlapping components  $E_1$ ,  $E_2$  and  $E_3$ . We simulate a sample of size  $n = 3000$ . Parameters of the model are given in Table 1. The three shape matrices

are the same for all the components and set to  $\begin{pmatrix} 1000^2 & 0 & 0 \\ 0 & 500^2 & 0 \\ 0 & 0 & 500^2 \end{pmatrix}$ .

Table 1: Exact parameters values.

	$\mu_x$	$\mu_y$	$\mu_z$	$\sigma$	$\pi$
E1	0	0	0	0.01	1/3
E2	1800	0	0	0.01	1/3
E3	3600	0	0	0.01	1/3

Results are presented on Figures 5, 6, 7 and 8. We observe that the variability of the estimations decreases when the sample size increases. We remark an important number of bad estimations

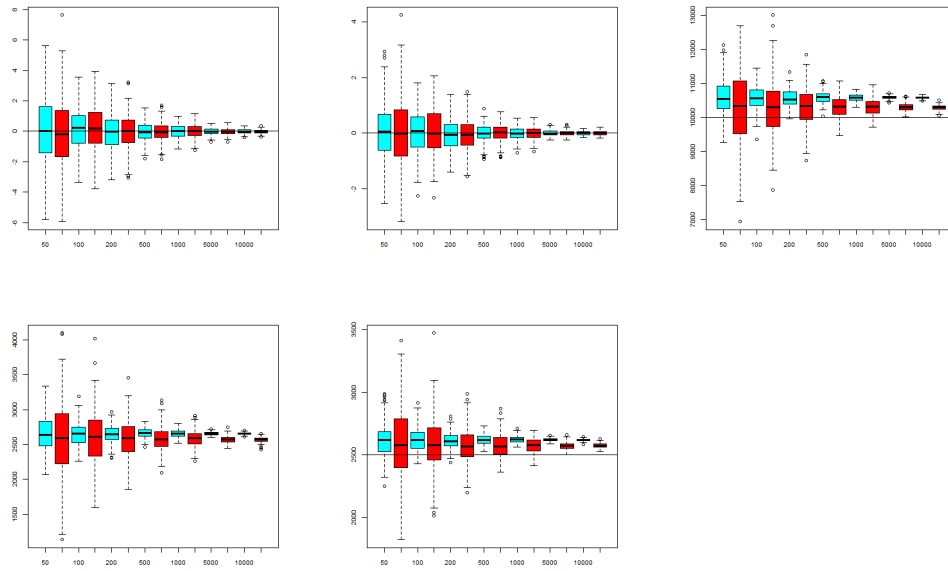


Figure 4: On the first row, boxplots of estimates of  $\mu_x$ ,  $\mu_y$  and  $\Sigma_{11}$ . On the second row, estimates of  $\Sigma_{12}$  and  $\Sigma_{22}$ . The BF method is plotted in red and the reference method is plotted in blue.

when the sample size is small. The estimation of the shape matrix is more difficult, especially for  $E_2$ , because of the intersection of the distributions. Observations located between two ellipsoids contributes to the estimation of the two component.

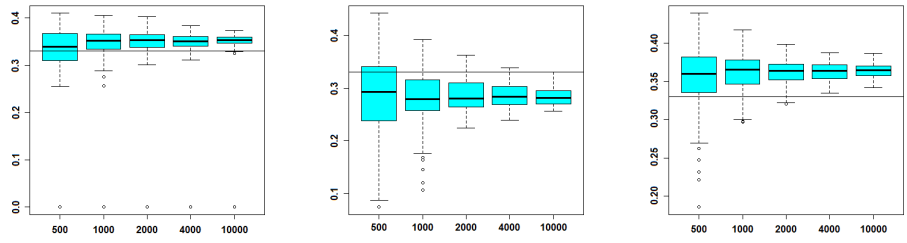


Figure 5: Boxplots of estimates of the mixture weights  $\pi_1$ ,  $\pi_2$  and  $\pi_3$ .

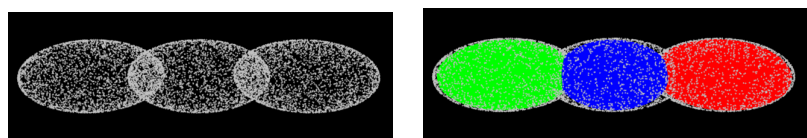


Figure 6: A sample of size  $n = 10000$  and the estimated model.

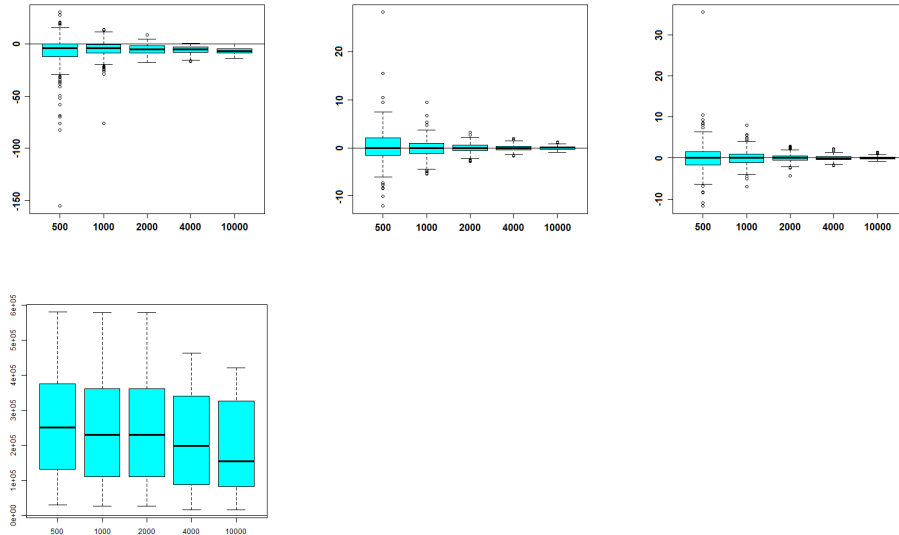


Figure 7: Estimates of  $E_1$  parameters. On the first row, boxplot of estimates of  $\mu_x$ ,  $\mu_y$  and  $\mu_z$ . On the second row, estimates of  $E(\Sigma)$ .

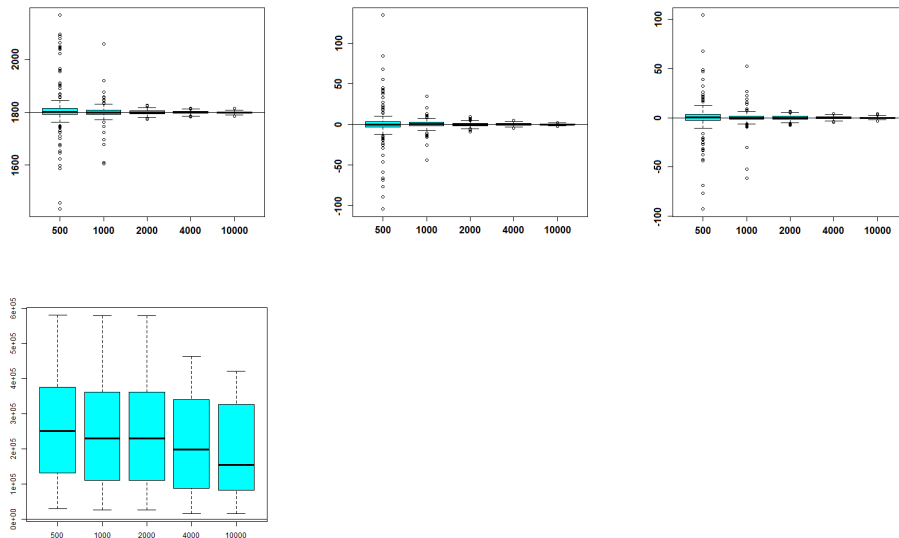


Figure 8: Estimates of  $E_2$  parameters. On the first row, boxplot of estimates of  $\mu_x$ ,  $\mu_y$  and  $\mu_z$ . On the second row, estimates of  $E(\Sigma)$ .

## 5 Applications

In this section, we experiment our ellipsoidal mixture model on real data. The objective is to model a person in 3D. Data are acquired with two calibrated depth sensor located on the opposite sides of a room. The two point clouds are fused thanks to a 3D registration technique [2]. The data are then segmented to keep only points belonging to the person. We apply the proposed algorithm with

$K = 2$  and  $K = 3$  components in the mixture model. Results are given in Figure 9. The model with  $K = 2$  components gives a better representation of the point cloud. The head is modelled by a nearly spherical shape while the body is modelled by a single ellipsoid. In the case of three components, the third ellipsoid is not meaningful.

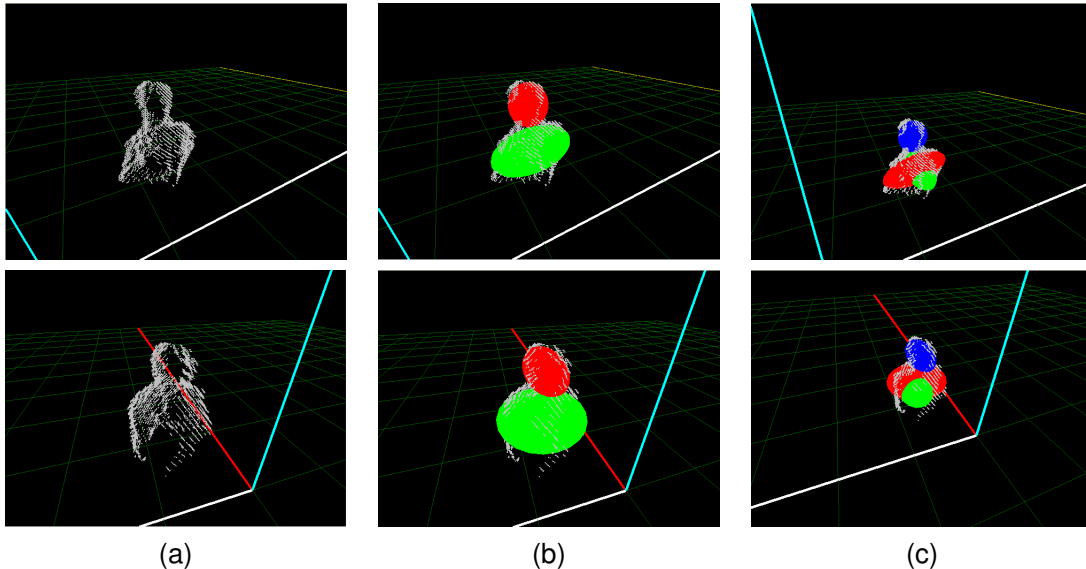


Figure 9: On the first column, 3D point clouds representing a standing person. On the second column, the estimated model with  $K = 2$  components, and on the third column, the estimated model with  $K = 3$  components.

## Conclusion

In this work, we proposed a new ellipsoidal mixture model for 3D data modelling. The mixture model is based on a new pdf belonging to the family of elliptical distributions and that models points spread around an ellipsoidal surface. The parameters of the mixture model based on this density are estimated with an EM based algorithm. The convergence of the method has been established theoretically. Our algorithm is simple to implement and produces accurate estimations. We obtained results comparable to a state of the art ellipse fitting method.

Experiments revealed that the analysis of the shape matrix  $\Sigma$  could be a promising way to determine if a given set of points is a sphere or an ellipsoid. An adequate statistical test could therefore be set up to classify objects.

## A Proofs and technical calculus

### A.1 Proofs of Lemma 2.1 and 2.2

*Proof of Lemma 2.1.* The proof is straightforward using the decomposition  $X = \mu + \Sigma^{1/2} W U$ . The random vector  $U$  is distributed on the  $d$ -dimensional unit sphere, which implies that  $\|U\| = 1$ ,  $\mathbb{E}[U] = 0$  and  $\mathbb{V}[U] = d^{-1} I_d$ . We therefore deduce that  $\mathbb{E}[X] = \mu$  and  $\mathbb{V}[X] = d^{-1} \mathbb{E}[W^2] \Sigma = d^{-1} \Sigma_*$ . We can conclude the first part of the proof by remarking that for  $q \geq 0$ ,  $\mathbb{E}[W^q] = J_{d+q-1} J_{d-1}^{-1}$ .

Then, to prove the second part, we set

$$\xi = \sqrt{(X - \mu)^T \Sigma_*^{-1} (X - \mu)},$$

and we can show that  $\xi = (\mathbb{E}[W^2])^{-1/2} W$ , leading to

$$\mathbb{V} \left( \sqrt{(X - \mu)^T \Sigma_*^{-1} (X - \mu)} \right) = \mathbb{E}(\xi^2) - (\mathbb{E}[\xi])^2 = 1 - \frac{(\mathbb{E}[W])^2}{\mathbb{E}[W^2]},$$

which closes the proof of the lemma.  $\square$

*Proof of Lemma 2.2.* The proof is based on the decomposition  $X = \mu + \Sigma^{1/2} W U$ . The random vector  $X^*$  can therefore be rewritten as  $X^* = \mu + \Sigma^{1/2} U (W - 1)$ . Then, since random variables  $W$  and  $U$  are independent with  $\mathbb{E}[U] = 0$ , we deduce that  $\mathbb{E}[X^*] = \mu$ .

This result enables us to compute the variance:  $\mathbb{V}[X^*] = \mathbb{E}[(X^* - \mu)(X^* - \mu)^T]$ . Given that variables  $W$  and  $U$  are independent with  $\mathbb{V}[U] = d^{-1} I_d$  and by remarking that  $\mathbb{E}[W^q] = J_{d+q-1} J_{d-1}^{-1}$ , we obtain  $\mathbb{V}[X^*] = d^{-1} \mathbb{E}[(W - \sqrt{\mathbb{E}[W^2]})^2] \Sigma$ .  $\square$

## A.2 Proof of Theorem 3.1

*Proof of Theorem 3.1.* Since  $\mathbb{E}[\|X_1\|^4] < \infty$ , using the law of the iterated logarithm of Hartman and Winters (see [36]), we straight forwardly obtain the first two results. We now focus on the convergence properties of the estimator  $\hat{\sigma}_n^2$  presented in Section 3.1. For this purpose, we first establish the convergence of  $\hat{\Sigma}_n^{-1}$  towards  $\Sigma_*^{-1}$ . The proof is standard and we resume the main steps of the proof proposed by N'Guyen and Saracco in [23].

From the Riccati equation for matrix inversion given for example in [8] (page 96), we can write the following decomposition

$$\hat{\Sigma}_n^{-1} = \Sigma_*^{-1} - \Sigma_*^{-1} (\hat{\Sigma}_n - \Sigma_*) \Sigma_*^{-1} + R_n,$$

where  $R_n$  is given by

$$R_n = \Sigma_*^{-1} (\Sigma_* - \hat{\Sigma}_n) \hat{\Sigma}_n^{-1} (\Sigma_* - \hat{\Sigma}_n) \Sigma_*^{-1}.$$

Then, we immediately deduce that

$$\left\| \hat{\Sigma}_n^{-1} - \Sigma_*^{-1} \right\| \leq \lambda_{\max}^2(\Sigma_*^{-1}) \left\| \hat{\Sigma}_n - \Sigma_* \right\| + \|R_n\|$$

and

$$\|R_n\| \leq \lambda_{\max}(\hat{\Sigma}_n^{-1}) \lambda_{\max}^2(\Sigma_*^{-1}) \left\| \Sigma_* - \hat{\Sigma}_n \right\|^2.$$

Let us recall that for any square matrix  $A$  positive definite, we have  $\lambda_{\max}(A^{-1}) = (\lambda_{\min}(A))^{-1}$ . Consequently, since  $\lambda_{\min}(\Sigma) > 0$  then  $\lambda_{\min}(\Sigma_*) > 0$ . In addition, since  $\hat{\Sigma}_n$  converges almost surely towards  $\Sigma_*$ , then  $\lambda_{\min}(\hat{\Sigma}_n) > 0$ . We therefore have

$$\|R_n\| = O \left( \left\| \Sigma_* - \hat{\Sigma}_n \right\|^2 \right) \quad \text{a.s.} \quad (\text{A.1})$$



Thanks to the upper bound  $\left\| \widehat{\Sigma}_n - \Sigma_* \right\| = O(a_n)$ , with  $a_n = \sqrt{(\log \log n)/n}$ , established in Section 3.1, we can write

$$\left\| \widehat{\Sigma}_n^{-1} - \Sigma_*^{-1} \right\| = O(a_n) \quad \text{a.s.} \quad (\text{A.2})$$

Let us now study the convergence of  $\widehat{\sigma}_n^2$  that can be rewritten under the form  $A_n - B_n^2$  with

$$A_n = \frac{1}{n} \sum_{j=1}^n \widehat{\xi}_j^2 \quad \text{and} \quad B_n = \frac{1}{n} \sum_{j=1}^n \widehat{\zeta}_j,$$

where we recall that  $\widehat{\zeta}_j = \sqrt{(X_j - \bar{X}_n)^T \widehat{\Sigma}_n^{-1} (X_j - \bar{X}_n)}$ . In order to study the convergence of  $A_n$ , it is rewritten under the form

$$A_n = A_{1,n} + A_{2,n} - (\bar{X}_n - \mu)^T \Sigma_*^{-1} (\bar{X}_n - \mu) \quad (\text{A.3})$$

with

$$\begin{aligned} A_{1,n} &= \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^T \left( \widehat{\Sigma}_n^{-1} - \Sigma_*^{-1} \right) (X_j - \bar{X}_n) \\ A_{2,n} &= \frac{1}{n} \sum_{j=1}^n (X_j - \mu)^T \Sigma_*^{-1} (X_j - \mu). \end{aligned}$$

We can easily show that  $\sum_{j=1}^n \|X_j - \bar{X}_n\|^2 = O(n)$  a.s., and by using (A.2), we obtain  $|A_{1,n}| = O(a_n)$  a.s. On the other hand, we have

$$A_{2,n} = \frac{1}{n} \sum_{j=1}^n \frac{J_2(\sigma)}{J_4(\sigma)} W_j^2 \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 1$$

and since the  $(W_j)$  have a finite moment of order 4, the law of the iterated logarithm of Hartman and Winters [36] can be applied, leading to  $|A_{2,n} - 1| = O(a_n)$  a.s.

Finally, as  $\|\bar{X}_n - \mu\| = O(a_n)$  a.s., we get from (A.3) the almost sure convergence of  $A_n$  towards 1 and the upper bound

$$|A_n - 1| = O(a_n) \quad \text{a.s.} \quad (\text{A.4})$$

Let us now study the convergence of  $B_n$ . To this aim,  $B_n$  is rewritten under the form

$$B_n = B_{1,n} + B_{2,n} + B_{3,n} \quad (\text{A.5})$$

with

$$\begin{aligned} B_{1,n} &= \frac{1}{n} \sum_{j=1}^n \tilde{\zeta}_j \\ B_{2,n} &= \frac{1}{n} \sum_{j=1}^n \left( \sqrt{(X_j - \mu)^T \widehat{\Sigma}_n^{-1} (X_j - \mu)} - \tilde{\zeta}_j \right) \\ B_{3,n} &= \frac{1}{n} \sum_{j=1}^n \left( \widehat{\zeta}_j - \sqrt{(X_j - \mu)^T \widehat{\Sigma}_n^{-1} (X_j - \mu)} \right) \end{aligned}$$

where we recall that  $\xi_j = \sqrt{(X_j - \mu)^T \Sigma_*^{-1} (X_j - \mu)}$ . We immediately deduce that for any  $1 \leq j \leq n$ ,

$$(\lambda_{\min}(\Sigma_*^{-1}))^{1/2} \|X_j - \mu\| \leq \xi_j \leq (\lambda_{\max}(\Sigma_*^{-1}))^{1/2} \|X_j - \mu\|.$$

In addition, by remarking that  $\xi_j$  can be written  $\xi_j = (\mathbb{E}[W^2])^{-1/2} W_j$ , we deduce that

$$B_{1,n} \xrightarrow[n \rightarrow \infty]{a.s.} \frac{\mathbb{E}[W]}{(\mathbb{E}[W^2])^{1/2}} \quad \text{and} \quad \left| B_{1,n} - \frac{\mathbb{E}[W]}{(\mathbb{E}[W^2])^{1/2}} \right| = O(a_n) \text{ a.s.} \quad (\text{A.6})$$

By using the equality  $\sqrt{a} - \sqrt{b} = (a - b)(\sqrt{a} + \sqrt{b})^{-1}$ , we easily obtain

$$\begin{aligned} |B_{2,n}| &\leq \frac{1}{n} \sum_{j=1}^n \frac{|(X_j - \mu)^T \widehat{\Sigma}_n^{-1} (X_j - \mu) - \xi_j^2|}{\sqrt{(X_j - \mu)^T \widehat{\Sigma}_n^{-1} (X_j - \mu) + \xi_j}} \\ &\leq \frac{1}{n} \sum_{j=1}^n \frac{|(X_j - \mu)^T (\widehat{\Sigma}_n^{-1} - \Sigma_*^{-1}) (X_j - \mu)|}{\xi_j}. \end{aligned}$$

Since  $\xi_j \geq (\lambda_{\min}(\Sigma_*^{-1}))^{1/2} \|X_j - \mu\|$ , we deduce that

$$|B_{2,n}| \leq (\lambda_{\max}(\Sigma_*))^{1/2} \left\| \widehat{\Sigma}_n^{-1} - \Sigma_*^{-1} \right\| \frac{1}{n} \sum_{j=1}^n \|X_j - \mu\|.$$

Finally, as  $\sum_{j=1}^n \|X_j - \mu\| = O(n)$  a.s., we have

$$|B_{2,n}| = O(a_n) \text{ a.s.} \quad (\text{A.7})$$

Thanks to the triangular inequality (derived from the Mahanobis distance), we get

$$|B_{3,n}| \leq \frac{1}{n} \sum_{j=1}^n \sqrt{(\bar{X}_n - \mu)^T \widehat{\Sigma}_n^{-1} (\bar{X}_n - \mu)} \leq \sqrt{\lambda_{\max}(\widehat{\Sigma}_n^{-1})} \|\bar{X}_n - \mu\|.$$

Finally, since on the one hand  $\|\bar{X}_n - \mu\| = O(a_n)$  a.s. and on the other hand,  $\widehat{\Sigma}_n^{-1}$  converges almost surely towards  $\Sigma_*^{-1}$  with  $\lambda_{\min}(\Sigma_*) > 0$ , we deduce that

$$|B_{3,n}| = O(a_n) \text{ a.s.} \quad (\text{A.8})$$

By combining (A.6), (A.7) and (A.8) with (A.5), we get

$$B_n \xrightarrow[n \rightarrow \infty]{a.s.} \frac{\mathbb{E}[W]}{(\mathbb{E}[W^2])^{1/2}} \quad \text{and} \quad \left| B_n - \frac{\mathbb{E}[W]}{(\mathbb{E}[W^2])^{1/2}} \right| = O(a_n) \text{ a.s.} \quad (\text{A.9})$$

The result on the convergence of  $\widehat{\sigma}_n^2 = A_n - B_n^2$  is obtained by combining the results (A.4) and (A.9). Indeed, we first obtain

$$\widehat{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{a.s.} 1 - \frac{(\mathbb{E}(W))^2}{\mathbb{E}(W^2)} = 1 - \frac{J_3^2}{J_2 J_4} = \widetilde{\sigma}^2.$$

Then, by considering the decomposition

$$\widehat{\sigma}_n^2 - \widetilde{\sigma}^2 = (A_n - 1) - (B_n - \gamma)(B_n + \gamma),$$

where we set  $\gamma = (\mathbb{E}[W^2])^{-1/2} \mathbb{E}[W]$ , and by using the upper bound  $|B_n + \gamma| = O(1)$  a.s., we deduce that

$$|\widehat{\sigma}_n^2 - \widetilde{\sigma}^2| = O(a_n) \text{ a.s.}$$

which closes the proof of the convergence of  $\widehat{\sigma}_n^2$ .  $\square$

## B Some important results on the normalization constant of the ellipsoidal density

This appendix deals with the computation of the normalization constant  $C_d$  of the new ellipsoidal density  $f$  introduced in (2.4). We first establish the following technical lemma which is useful to simplify computations.

**Lemma B.1.** *Let us set for  $q \in \mathbb{N}$  and  $\alpha > 0$ ,*

$$J_q(\alpha) = \int_0^\infty t^q \exp(-(t-1)^2 / (2\alpha^2)) dt. \quad (\text{B.1})$$

Then  $J_0(\alpha) = \alpha\sqrt{2\pi}(1 - \Phi(-1/\alpha))$ ,  $J_1(\alpha) = J_0(\alpha) + \alpha^2 \exp(-1/(2\alpha^2))$ , and for  $q \geq 2$ ,

$$J_q(\alpha) = J_{q-1}(\alpha) + (q-1)\alpha^2 J_{q-2}(\alpha) \quad (\text{B.2})$$

where  $\Phi$  denotes the gaussian cumulative distribution function.

*Proof.* Let us set  $v(t) = -(t-1)^2 / (2\alpha^2)$ . Then,  $v'(t) = -(t-1)/\alpha^2$  and we can rewrite (B.1) under the form

$$\begin{aligned} J_q(\alpha) &= \int_0^\infty (t^q - t^{q-1}) \exp(v(t)) dt + J_{q-1}(\alpha) \\ &= -\alpha^2 \int_0^\infty t^{q-1} v'(t) \exp(v(t)) dt + J_{q-1}(\alpha) \end{aligned} \quad (\text{B.3})$$

Finally, (B.2) immediately follows using an integration by parts. To close the proof, it is clear that  $J_0(\alpha) = \alpha\sqrt{2\pi}(1 - \Phi(-1/\alpha))$  and we get  $J_1(\alpha)$  using (B.3) with  $d = 1$ .  $\square$

Let us now consider the constant  $C_d$  which satisfies

$$I = C_d \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2\sigma^2} \left(\sqrt{(x-\mu)^T \Sigma^{-1} (x-\mu)} - 1\right)^2\right) dx = 1.$$

The substitution  $y := \Sigma^{-1/2} (x - \mu)$  leads to

$$C_d \int_{\mathbb{R}^d} |\Sigma^{1/2}| \exp\left(-\frac{1}{2\sigma^2} (\|y\| - 1)^2\right) dy = 1.$$

Then, a polar coordinates substitution leads to

$$\begin{aligned} I &= C_d |\Sigma|^{1/2} \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^{+\infty} \rho^{d-1} \exp\left(-\frac{(\rho-1)^2}{2\sigma^2}\right) d\rho \\ &= C_d |\Sigma|^{1/2} \frac{2\pi^{d/2}}{\Gamma(d/2)} J_{d-1}(\sigma) = 1. \end{aligned}$$

The expression of the normalization constant is

$$C_d = \frac{\Gamma(d/2)}{2\pi^{d/2} |\Sigma|^{1/2} J_{d-1}(\sigma)}.$$

For the first values of  $d$ , the expressions of the constant are given by

$$C_1 = \frac{1}{2\sigma\sqrt{2\pi} |\Sigma|^{1/2} [1 - \Phi(-1/\sigma)]}$$

$$C_2 = \frac{1}{(2\pi)^{3/2} |\Sigma|^{1/2} \sigma [1 - \Phi(-1/\sigma)]}$$

$$C_3 = \frac{1}{2(2\pi)^{3/2} |\Sigma|^{1/2} \sigma(1 + \sigma^2) \left[ 1 - \Phi(-1/\sigma) + \frac{\sigma \exp(-1/(2\sigma^2))}{(1 + \sigma^2)\sqrt{2\pi}} \right]}$$

where  $\Phi$  denotes the gaussian cumulative distribution function.

We observe that the parameter  $\sigma$  plays a particular role in these expressions. Indeed, under the assumption that the parameter  $\sigma$  is small enough, terms  $\Phi(-1/\sigma)$  and  $\frac{t}{\sqrt{2\pi(1+\sigma^2)}} e^{(-1/(2\sigma^2))}$  can be ignored (cf Appendix C). In that case, we can approximate the normalization constant and thus obtain

$$C_1 = \frac{1}{2\sigma\sqrt{2\pi} |\Sigma|^{1/2}}$$

$$C_2 = \frac{1}{(2\pi)^{3/2} |\Sigma|^{1/2} \sigma}$$

$$C_3 = \frac{1}{2(2\pi)^{3/2} |\Sigma|^{1/2} \sigma(1 + \sigma^2)}.$$

## C Influence of parameter $\sigma^2$

This appendix concerns the influence of parameter  $\sigma^2$  on several quantities.

Figure 10 represents the values of terms  $\Phi(-1/\sigma)$  and  $\frac{t}{\sqrt{2\pi(1+\sigma^2)}} e^{(-1/(2\sigma^2))}$  with respect to  $\sigma$ . We observe that these quantities are close to zero for small values of  $\sigma$ .

Figure 11 shows the values of terms  $\frac{1 + 6\sigma^2 + 3\sigma^4}{1 + \sigma^2}$ ,  $\frac{1 + 3\sigma^4}{1 + 7\sigma^2 + 9\sigma^4 + 3\sigma^6}$  and  $\frac{1 + 3\sigma^2}{1 + \sigma^2}$  with respect to  $\sigma$ . We can therefore quantify the difference between the given terms and the value 1.

## D Justification of the M-step in the EM algorithm

In this appendix, we justify the choice of the estimators used in the M-step of the EM algorithm described in Section 3.3. To maximize the conditional expectation of the completed log-likelihood of parameters  $\Theta = (\pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K)$  with  $\theta_k = (\mu_k, \Sigma_k, \sigma_k)$ , we compute partial derivatives with respect to each parameter of the quantity

$$Q(\Theta) = \sum_{i=1}^n \sum_{k=1}^K t_{ik} (\log(\pi_k) + \log(f(x_i, \theta_k))),$$

where

$$f(x | \theta_k) = \frac{(2\pi)^{-3/2} |\Sigma_k|^{-1/2}}{2\sigma_k(1 + \sigma_k^2)} \exp\left(-\frac{1}{2\sigma_k^2} \left(\sqrt{(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)} - 1\right)^2\right)$$

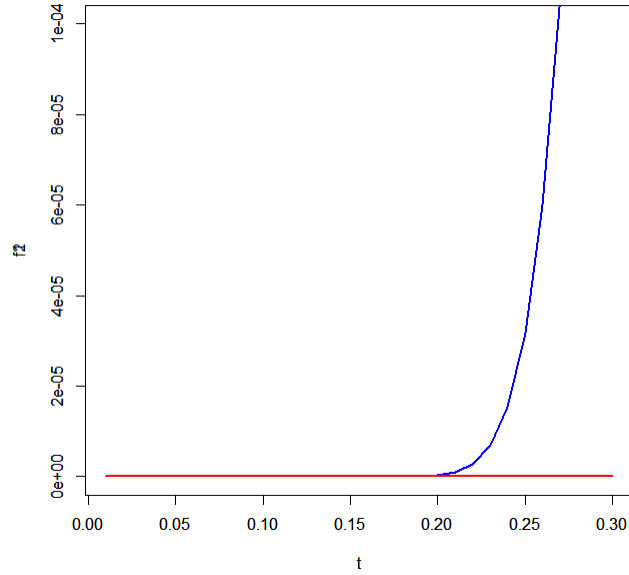


Figure 10: Values of terms  $\Phi(-1/\sigma)$  (blue) and  $\frac{t}{\sqrt{2\pi(1+\sigma^2)}}e^{(-1/(2\sigma^2))}$  (red) with respect to  $\sigma$ .

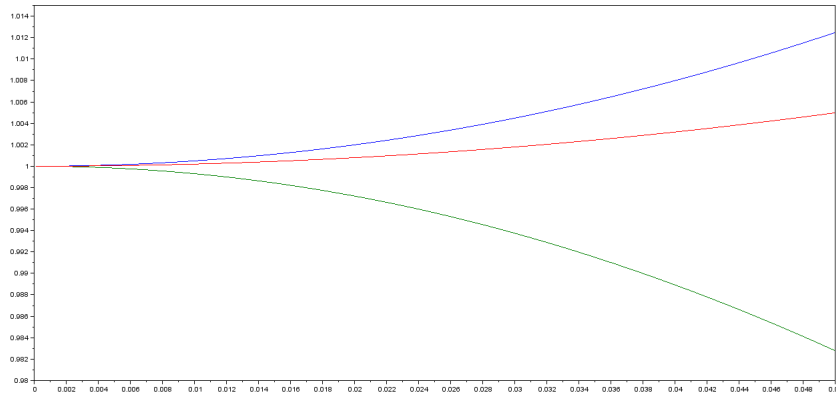


Figure 11: Values of terms  $\frac{1 + 6\sigma^2 + 3\sigma^4}{1 + \sigma^2}$  (blue),  $\frac{1 + 3\sigma^4}{1 + 7\sigma^2 + 9\sigma^4 + 3\sigma^6}$  (green) and  $\frac{1 + 3\sigma^2}{1 + \sigma^2}$  (red) with respect to  $\sigma$ .

and with  $\sigma$  sufficiently small. The partial derivative of  $Q$  with respect to  $\mu_k$  is

$$\frac{\partial Q}{\partial \mu_k} = \frac{\Sigma_k^{-1}}{\sigma_k^2} \sum_{i=1}^n t_{ik} (x_i - \mu_k) \left[ 1 - \frac{1}{\sqrt{(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)}} \right].$$

The partial derivative of  $Q$  with respect to  $\Sigma_k^{-1}$  is

$$\frac{\partial Q}{\partial \Sigma_k^{-1}} = \sigma_k^2 \Sigma_k - \frac{1}{n} \sum_{i=1}^n t_{ik} (x_i - \mu_k)(x_i - \mu_k)^T \left[ 1 - \frac{1}{\sqrt{(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)}} \right].$$

The partial derivative of  $Q$  with respect to  $\sigma_k^2$  is

$$\frac{\partial Q}{\partial \sigma_k} = -n \frac{(1 + 3\sigma_k^2)}{\sigma_k(1 + \sigma_k^2)} + \frac{1}{\sigma_k^3} \sum_{i=1}^n t_{ik} \left( \sqrt{(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)} - 1 \right)^2. \quad (\text{D.1})$$

The maximization problem cannot be solved analytically. Under the assumption that  $\sigma_k$  is small enough, we can write  $(1 + 3\sigma_k^2) / (1 + \sigma_k^2) \simeq 1$  and the partial derivative (D.1) can be simplified. However, for sake of simplicity, we prefer estimate the shape matrix  $\Sigma_k$  by the usual estimator  $n^{-1} \sum_{i=1}^n (x_i - \mu_k)(x_i - \mu_k)^T$ . This choice is also motivated by our experiments.

We can easily show that the mixture weights  $\{w_{ik}\}_{i=1..n}^{k=1..K}$  can be estimated using the classical estimators employed for gaussian mixture models. Finally, these computations lead to the estimators presented in Section 3.3.

## References

- [1] Sung Joon Ahn, Wolfgang Rauh, Hyung Suck Cho, and H-J Warnecke Orthogonal distance fitting of implicit curves and surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):620–638, 2002.
- [2] Paul J. Besl. A Method for Registration of 3-D Shape. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Los Alamitos, CA, USA, vol. 14, no 2, 1992, p. 239–256.
- [3] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer.
- [4] D. Brazey and B. Portier (2014). A new spherical mixture model for head detection in depth images. *SIAM Journal on Imaging Science*, Vol. 7, No. 4, pp. 2423-2447.
- [5] Burt, A, Disney, M, Calders, K. Extracting individual trees from lidar point clouds using treeseg. *Methods Ecol Evol*, 10: 438– 445, 2019.
- [6] A. P. Dempster, N. M. Laird and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, Vol. 39, pp. 1-38, 1977.
- [7] X. Du, M. H. Ang, S. Karaman and D. Rus. A General Pipeline for 3D Detection of Vehicles. *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3194-3200, 2018.
- [8] M. Duflo. *Random Iterative Models*, Springer Verlag, Berlin.
- [9] M. A. Fischler and R. C. Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis. *Communications of the ACM*, vol. 24, pp. 381-395.
- [10] Andrew W. Fitzgibbon and R.B.Fisher. A Buyer's Guide to Conic Fitting. *Proc.5th British Machine Vision Conference*, Birmingham, pp. 513-522.

- [11] N. Greggio, A. Bernardino, C. Laschi, P. Dario and J. Santos-Victor. Fast Estimation of Gaussian Mixture Models for Image Segmentation. *Machine Vision and Applications*, Vol. 23, pp. 773-789.
- [12] B. Jian and B. C. Vemuri. A Robust Algorithm for Point Set Registration Using Mixture of Gaussians. *International Conference on Computer Vision*, Vol. 2, pp. 1246-1251.
- [13] Martti Kesäniemi and Kai Virtanen Direct least square fitting of hyperellipsoids. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1):63–76, 2017.
- [14] Qingde Li and John G Griffiths Least squares ellipsoid specific fitting. *Geometric Modeling and Processing*, pages 335–340, 2004.
- [15] Li, L., Sung, M., Dubrovina, A., Yi, L. and Guibas, L. J. Supervised fitting of geometric primitives to 3d point clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2652-2660, 2019.
- [16] Li Y, Li W, Darwish W, Tang S, Hu Y, Chen W. Improving Plane Fitting Accuracy with Rigorous Error Models of Structured Light-Based RGB-D Sensors. *Remote Sensing*, 12(2):320, 2020.
- [17] J. Liu and Z. Wu. An adaptive approach for primitive shape extraction from point clouds. *Optik - International Journal for Light and Electron Optics*, 125(9), pages 2000-2008.
- [18] D.A. Martins, A.J. Neves and A.J. Pinho. Real-time generic ball recognition in RoboCup domain. *Proc. of the 3rd International Workshop on Intelligent Robotics*, pages 37-48.
- [19] G. J. McLachlan and D. Peel. Finite Mixture Models. *Wiley Series in Probability and Statistics*.
- [20] G. J. McLachlan and K. Thriyambakam. The EM Algorithm and Extensions. *WILEY-Interscience*.
- [21] R. J. Muirhead. Aspects of multivariate statistical theory, Wiley.
- [22] M. Nahangi, T. Czerniawski, Carl T. Haas, S. Walbridge. Pipe radius estimation using Kinect range cameras. *Automation in Construction*, vol. 99, pp. 197-205, 2019.
- [23] T.M.N. Nguyen and J. Saracco. Estimation récursive en régression inverse par tranches (sliced inverse regression). *Journal de la Société Française de Statistique*, vol. 151 pp. 19-46.
- [24] P. Núñez, P. Drews Jr, R. Rocha, M. Campos and J. Dias. Novelty detection and 3D shape retrieval based on gaussian mixture models for autonomous surveillance robotics,. *International Conference on Intelligent Robots and Systems*, pp. 4724-4730.
- [25] OpenCV library. <http://opencv.org/>.
- [26] Parr B., Legg M. Alam, F. Analysis of Depth Cameras for Proximal Sensing of Grapes. *Sensors*, no. 11: 4179, 2022.
- [27] Qi, C. R., Yi, L., Su, H., and Guibas, L. J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, vol. 30, 2017.
- [28] T. Rabbani. Automatic reconstruction of industrial installations using point clouds and images. *NCG Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission*.

- [29] C. Romanengo, A. Raffo, Y. Qie, N. Anwer, B. Falcidieno. Fit4CAD: A point cloud benchmark for fitting simple geometric primitives in CAD objects. *Computers and Graphics*, vol. 102, pp 133-143, 2022.
- [30] M.S. Shafiq, S.T. Tümer, S. Turgut and H.C. Güler. Marker detection and trajectory generation algorithms for a multicamera based gait analysis system,. *Mechatronics*, 11(4), pages 409-437.
- [31] Sharma, G., Liu, D., Maji, S., Kalogerakis, E., Chaudhuri, S., Měch, R. PARSENET: A Parametric Surface Fitting Network for 3D Point Clouds. *European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science*, pp 261–276, 2020.
- [32] M. Han, J. Kan and Y. Wang. Ellipsoid Fitting Using Variable Sample Consensus and Two-Ellipsoid-Bounding-Counting for Locating Lingwu Long Jujubes in a Natural Environment. *IEEE Access*, vol. 7, pp. 164374-164385, 2019.
- [33] Shi Y, Zhao G, Wang M, Xu Y, Zhu D. An Algorithm for Fitting Sphere Target of Terrestrial LiDAR. *Sensors*,21(22):7546, 2021.
- [34] C. Sommer, Y. Sun, E. Bylow, and D. Cremers. PrimiTect: Fast Continuous Hough Voting for Primitive Detection. *ICRA*, 2020.
- [35] C. Stauffer, E. Grimson. Adaptive background mixture models for real-time tracking. *Computer Vision and Pattern Recognition*.
- [36] W.F. Stout. Almost sure convergence. *A Series of Monographs and Textbooks in Probability and Mathematical Statistics*.
- [37] B. Sun and P. Mordohai. Oriented point sampling for plane detection in unorganized point clouds. *ICRA*,21(22), 2019.
- [38] A. Thom,. A statistical examination of the megalithic sites in Britain. *Journal of the Royal Statistical Society, Series A (General)*, pages 275-295.
- [39] Yang, B., Liu, C., Li, B., Jiao, J., Ye, Q. Prototype Mixture Models for Few-Shot Semantic Segmentation. *ECCV*, vol 12353, 2020.
- [40] X. Zhang, J. Stockel, M. Wolf, P. Cathier, G. McLennan, E.A. Hoffman and M. Sonka. A New Method for Spherical Object Detection and Its Application to Computer Aided Detection of Pulmonary Nodules in CT Images. *Medical Image Computing and Computer-Assisted Intervention*, vol. 4791, pages 842-849.
- [41] Zhao, Mingyang and Jia, Xiaohong and Ma, Lei and Qiu, Xinlin and Jiang, Xin and Yan, Dong-Ming. Robust Ellipsoid-specific Fitting via Expectation Maximization. *Conference BMVC*, 2021.
- [42] J. Zuo, Z. Jia, J. Yang and N. Kasabov Moving Target Detection Based on Improved Gaussian Mixture Background Subtraction in Video Images. *IEEE Access*, vol. 7, pp. 152612-152623, 2019.